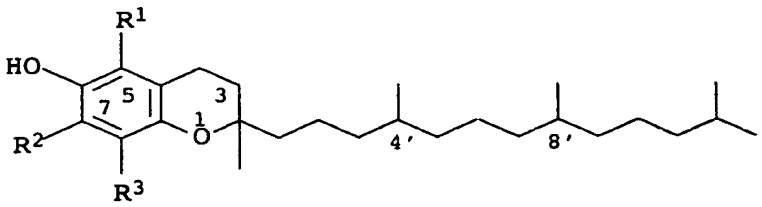
DNA sequence encoding a hydroxyphenylpyruvate dioxygenase, and its overproduction in plants

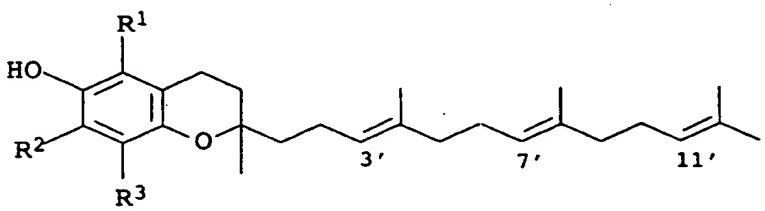
The present invention relates to a method of generating plants with an elevated vitamin E content by expressing an exogenous or endogenous HPPD gene in plants or plant organs. The invention furthermore relates to the use of the corresponding nucleic acids encoding an HPPD gene in transgenic plants to make the latter resistant to HPPD inhibitors, and to the use of the DNA sequence encoding an HPPD for generating a test system for identifying HPPD inhibitors.

An important aim in plant molecular genetics is the generation of plants with an elevated content of sugars, enzymes and amino acids. It would also be economically interesting to develop plants with an elevated vitamin content, eg. an elevated vitamin E content.

The eight naturally occurring compounds with vitamin E activity are derivatives of 6-chromanol (Ullmann's Encyclopedia of Industrial Chemistry, Vol. A 27 (1996), VCH Verlagsgesellschaft, Chapter 4., 478-488, Vitamin E). The first group (1a - d) is derived from tocol, while the second group is composed of tocotrienol derivatives (2a - d):



1a, α-tocopherol: $R^1 = R^2 = R^3 = CH_3$
1b, β-tocopherol [148-03-8]: $R^1 = R^3 = CH_3$, $R^2 = H$
1c, γ-tocopherol [54-28-4]: $R^1 = H$, $R^2 = R^3 = CH_3$
1d, δ-tocopherol [119-13-1]: $R^1 = R^2 = H$, $R^3 = CH_3$



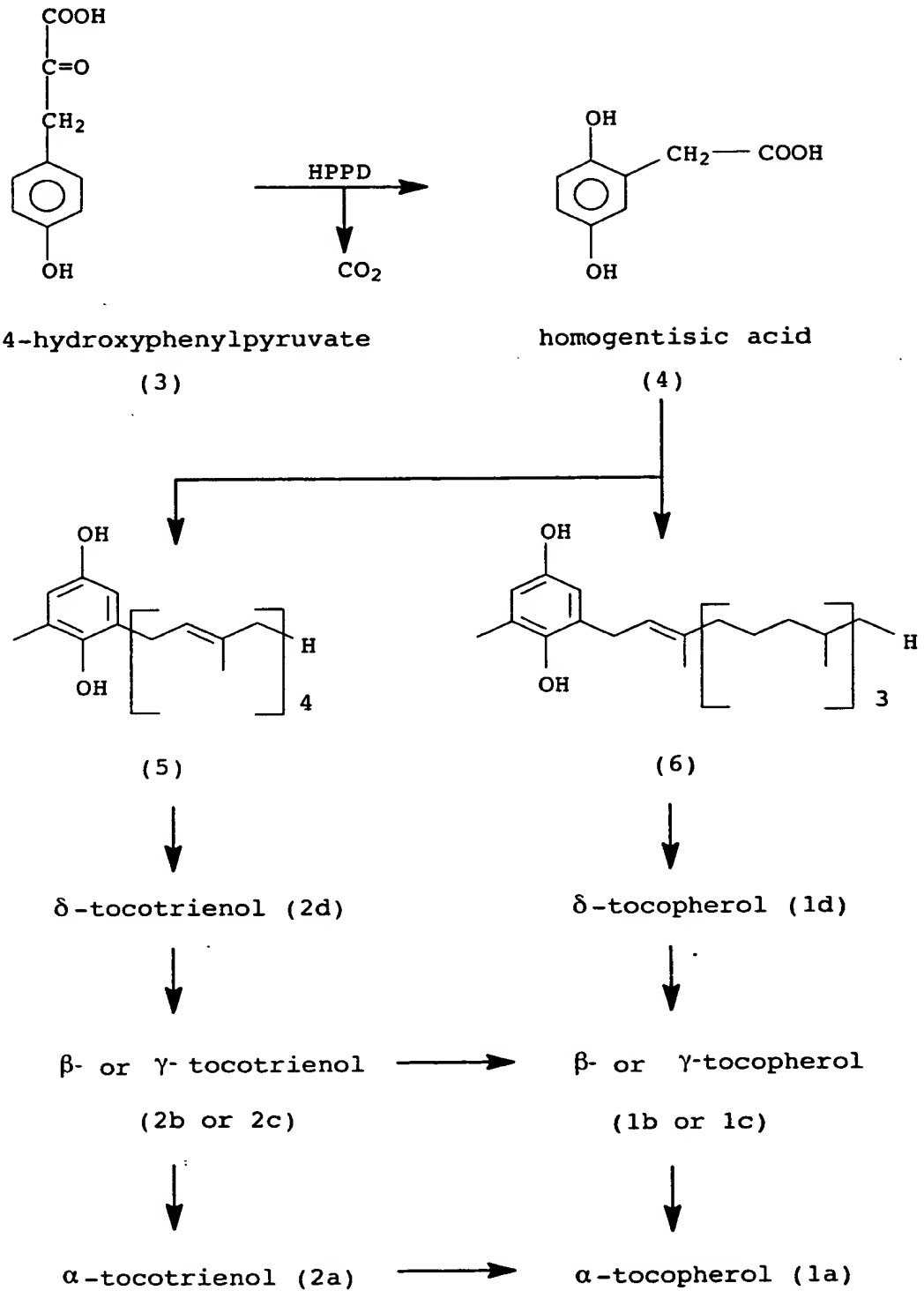
2a, α-tocotrienol [1721-51-3]: $R^1 = R^2 = R^3 = CH_3$
2b, β-tocotrienol [490-23-3]: $R^1 = R^3 = CH_3$, $R^2 = H$
2c, γ-tocotrienol [14101-61-2]: $R^1 = H$, $R^2 = R^3 = CH_3$
2d, δ-tocotrienol [25612-59-3]: $R^1 = R^2 = H$, $R^3 = CH_3$

α-Tocopherol is of great economic importance.

The development of crop plants with an elevated vitamin E content by means of tissue culture or seed mutagenesis and natural selection has its limits. On the one hand, the vitamin E content must be detectable as early as at the tissue culture level and, on the other hand, only those plants can be manipulated via tissue culture techniques which can successfully be regenerated into entire plants, starting from cell cultures. Moreover, following mutagenesis and selection, crop plants may show undesirable characteristics which have to be eliminated by back-crossing, in some cases repeated back-crossing. Also, elevation of the vitamin E content by means of crossing would be limited to plants of the same species.

Those are the reasons why the genetic engineering approach, viz. isolating an essential biosynthesis gene which encodes the vitamin E synthesis performance and transferring it specifically into crop plants, is superior to the traditional breeding method. The conditions for this method are that the biosynthesis and its regulation are known and that genes which affect biosynthesis performance are identified.

Tocopherol biosynthesis in plants and algae proceeds in a known manner and is as follows:


COOH
C=O
CH2
OH
HPPD
CO2
OH
CH2 — COOH
OH
4-hydroxyphenylpyruvate
(3)
homogentisic acid
(4)
OH
OH
H
4
(5)
OH
OH
H
3
(6)
δ-tocotrienol (2d)
δ-tocopherol (1d)
β- or γ- tocotrienol
(2b or 2c)
β- or γ-tocopherol
(1b or 1c)
α-tocotrienol (2a)
α-tocopherol (1a)

The precursor of the aromatic ring of the tocopherols is p-hydroxyphenylpyruvate (3), which is converted enzymatically into homogentisic acid (4) with the aid of the enzyme hydroxyphenylpyruvate dioxygenase (HPPD), and the homogentisic acid reacts with phytyl pyrophosphate with elimination of CO_2 to give the precursor (6). The tocotrienol biosynthesis route starts with a condensation reaction between homogentisic acid (4) and geranylgeranyl pyrophosphate to give the precursor (5). Enzymatic cyclization of the precursors 5 or 6 gives δ- tocotrienol or δ-tocopherol, respectively. Some of these biosynthesis enzymes have been isolated.

While searching for Arabidopsis mutants with defects in the carotinoid biosynthesis, a white phenotype mutant was identified which is not capable of producing active HPPD. If this mutant, termed *pds2*, is raised in the presence of homogentisic acid, it produces carotinoids, like the wild type, and greens (Norris et al., Plant Cell (1995) 7: 2139 - 2149). This work shows that HPPD activity is a prerequisite for the formation of photosynthetically active chloroplasts. Without this enzyme, no plastoquinones are formed, which are required as acceptors for liberated reduction equivalents during carotinoid biosynthesis (phytoene desaturation). The fact that HPPD has a key role in the plastid metabolism makes it an interesting target for herbicides. Sulcotriones efficiently inhibit the activity of the enzyme (Schultz et al., FEBS Lett. (1993) 318: 162 - 166).

Sequences of HPPD-specific genes are already known from the organisms mentioned below:

Organism	Sequence name	Access number database
Humans	HPPD_HUMAN	X72389
Pig	HPPD_PIG	D13390
Rat	HPPD_RAT	M18405
Mouse	HPPD_MOUSE	D29987
Streptomyces avermitilis	SA11864	U11864
Pseudomonas sp. strain P.J. 874	HPPD_PSESP	P80064
Arabidopsis	HPPD_ARAB1	AF900228
	HPPD_ARAB2	U89267

Furthermore, the following sequences, which show a marked homology with HPPD sequences, can be found in the databases:

PEA3_MOUSE: Mus muscula (mouse) PEA3 polypeptide, AC X63190;

MELA_SHECO: Shewanella colwelliana, melA protein, AC M59289.

WO 96/38567 describes the HPPD DNA sequence from *Arabidopsis thaliana* and *Daucus carota*.

A knowledge of the HPPD DNA sequences is an absolute prerequisite both for the use in crop protection for the generation of herbicide-resistant plants and for increasing the vitamin E synthesis in plants, for example for producing animal feeds with elevated vitamin E content.

It is an object of the present invention to develop a transgenic plant with elevated vitamin E content.

It is a further object of the present invention to develop a transgenic plant which is resistant to HPPD inhibitors.

We have found that these objects are achieved, surprisingly, by overexpressing an HPPD gene in the plants.

It is an additional object of the present invention to develop a test system for identifying HPPD inhibitors.

We have found that this object is achieved by expressing a barley HPPD gene in a plant or in a microorganism and subsequently testing chemicals for inhibition of HPPD enzyme activity.

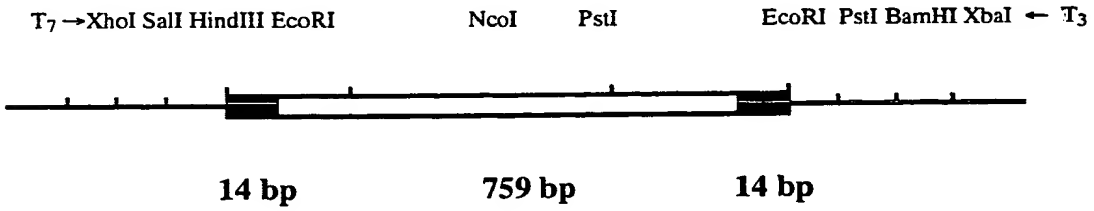
A first aspect of the present invention relates to the cloning of the complete barley HPPD gene via isolating the HPPD-gene-specific cDNA (HvSD36).

During leaf senescence, the vitamin E content in the leaves is markedly increased (Rise et al., Plant Physiol. (1989) 89: 1028 - 1030). The monocotyledonous leaf of barley represents a gradient of cells of different ages since the leaf has a basal meristem, from which new cells are formed by successive division. Thus, the oldest cells are located at the leaf tip and the youngest at the base. Fig. 1 shows a diagram of the primary leaf of barley on various days after sowing. The total leaf length measured can be seen from the scale on the left-hand side. Shown, and termed I - IV, are the leaf sections of the primary leaf which are differentiated to various degrees and which have been

selected for gene expression analysis. The plants were raised in a daily light/dark photoperiod (L/D) and, for inducing senescence, were excised after 6 days and incubated for 2 days in the dark (2 nD). A "Northern blot" analysis of RNA from the barley primary leaf from sections which had differentiated to various degrees (see Fig. 2) suggest that HPPD expression in barley is controlled in a development-dependent manner. Thus, copious accumulation of the approx. 1600 nt long transcript takes place in the meristematic region on the primary leaf base (I). The content of this transcript decreases with increasing age of the tissue (IIa and IIb) and increases again in the fully differentiated cells with mature chloroplasts (III). Finally, the content of the 1600 nt long transcript is highest in the senescing sections of the primary leaf (IV). In addition, an approx. 3100 nt long transcript can be detected only in the meristematic cells on the base of the primary leaf. Again, this transcript can no longer be detected with increasing tissue maturation.

With the aid of the so-called "Differential Display" method, a 207 bp cDNA fragment was first isolated whose corresponding transcript accumulates in the primary leaf of barley in the case of dark-induced senescence. This fragment (sequence protocol: sequence ID NO:1: nucleotide position 1342 - 1549) was subsequently used as a probe to isolate a cDNA clone with a larger insert in a cDNA library (in λ-ZAP-II) from senescing barley flag leaves.

Diagram of the cDNA subclone HvSD 36 from the λ-ZAP-II library:



The cDNA fragment (sequence protocol: Sequence ID NO:1: nucleotide position 771 - 1529) was cloned into the EcoR1 cleavage site of pBluescript(SK⁻). In addition, both ends of the cDNA are equipped with a 14 bp adaptor sequence which was required for ligation into λ-ZAP-II. Selected restriction sites of the vector and of the cDNA itself are shown.

The 759 bp long cDNA fragment was used as probe in a further experiment to obtain a complete sequence of HvSD 36. To this end, a cDNA library from RNA of the meristematic section of 5-day-old

barley seedlings was available. The lambda phage ExCell Eco RICIP from Pharmacia (Freiburg) (product number: 27-5011, 45.5kb) was used for this cDNA library.

A 1565 bp long cDNA clone was isolated, see sequence protocol: sequence ID NO:1: and 2.

Amongst the sequences in the databases, the 434 amino acids long protein sequence has a homology of 58%, which is the highest homology with the HPPD sequence from *Arabidopsis thaliana*.

To find a genomic clone which contains the complete HPPD gene sequence, a lambda FIXII library of barley was obtained from Stratagene (Heidelberg, product number 946104). The library was prepared using DNA from etiolated leaves of winter barley cv. Igri. The DNA was subjected to partial digestion with Sau3AI. Prior to cloning into the Xho1 cleavage site of the vector, the fragment ends and the phage arms were filled up with nucleotides. Screening of the library with 200,000 pfu in the first round gave only one clone which hybridized with cDNA HvSD36. After subjecting this recombinant phage to restriction digestion with PstI and SacI, fragments of a size of 5400, 3800 and 1800 bp, respectively, were isolated which can be detected in a "Southern" blot hybridization with the HvSD36 probe. These sub-fragments exist in cloned form in the Bluescript vector. Figure 3 shows the construction of the barley HPPD gene in the form of a diagram.

The invention relates in particular to expression cassettes whose sequence encodes an HPPD or a functional equivalent thereof, and to the use of these expression cassettes for generating a plant with an elevated vitamin E content. The nucleic acid sequence may be, for example, a DNA or a cDNA sequence. Encoding sequences which are suitable for insertion into an expression cassette according to the invention are, for example, those which encode an HPPD and which impart, to the host, the ability to overproduce vitamin E.

In addition, the expression cassettes according to the invention comprise regulatory nucleic acid sequences which govern expression of the encoding sequence in the host cell. In accordance with a preferred embodiment, an expression cassette according to the invention comprises upstream, ie. on the 5' end of the encoding sequence, a promoter and downstream, ie. on the 3' end, a polyadenylation signal and, if appropriate, other regulatory elements which are operatively linked with the encoding sequence for the HPPD gene which is located in-between. Operative linkage is to be understood as meaning the sequential

arrangement of promoter, encoding sequence, terminator and, if appropriate, other regulatory elements in such a way that each of the regulatory elements can fulfill its function as intended when the encoding sequence is expressed. The sequences preferred for operative linkage, but not limited thereto, are targeting sequences for guaranteeing subcellular localization in the apoplast, in the vacuole, in plastids, in the mitochondrium, in the endoplasmatic reticulum (ER), in the nucleus, in liposomes or in other compartments and translation enhancers such as the 5' leader sequence from the tobacco mosaic virus (Gallie et al., Nucl. Acids Res. 15 (1987) 8693 - 8711).

For example, the plant expression cassette can be incorporated into the tobacco transformation vector pBinAR-Hyg. Fig. 4 shows the tobacco transformation vectors pBinAR-Hyg with 35S promoter (A) and pBinAR-Hyg with seed-specific promoter phaseolin 796 (B):

- HPT: hygromycin phosphotransferase
- OCS: octopine synthase terminator
- PNOS: nopaline synthase promoter
- those restriction sites which cleave the vector only once are also shown.

Suitable as promoters of the expression cassette according to the invention are, in principle, all promoters which can control the expression of foreign genes in plants. In particular a plant promoter or a promoter derived from a plant virus is preferably used. Particularly preferred is the CaMV 35S promoter from cauliflower mosaic viruss (Franck et al., Cell 21 (1980) 285 - 294). It is known that this promoter contains various recognition sequences for transcriptional effectors which in their entirety lead to permanent and constitutive expression of the gene introduced (Benfey et al., EMBO J. 8 (1989) 2195 - 2202).

The expression cassette according to the invention may additionally comprise a chemically inducible promoter by means of which expression of the exogenous HPPD gene in the plant can be controlled at a specific point in time. Such promoters which can be used are, inter alia, for example the PRP1 promoter (Ward et al., Plant. Mol. Biol. 22 (1993), 361-366), a promoter which can be induced by salicylic acid (WO 95/19443), a promoter which can be induced by benzenesulfonamide (EP-A 388186), a promoter which can be induced by tetracyclin (Gatz et al., (1992) Plant J. 2, 397-404), a promoter which can be induced by abscisic acid

(EP-A 335528) or a promoter which can be induced by ethanol or cyclohexanone (WO 93/21334).

Furthermore, particularly preferred promoters are those which ensure expression in tissues or plant organs in which the biosynthesis of vitamin E, or its precursors, takes place. Promoters which must be mentioned in particular are those which guarantee leaf-specific expression. Promoters which may be mentioned are the potato cytosolic FBPase or the potato ST-LSI promoter (Stockhaus et al., EMBO J. 8 (1989) 2445 - 245).

With the aid of a seed-specific promoter, it was possible stably to express a foreign protein in the seeds of transgenic tobacco plants in an amount of up to 0.67% of the total soluble seed protein (Fiedler and Conrad, Bio/Technology 10 (1995), 1090-1094). The expression cassette according to the invention can therefore contain, for example, a seed-specific promoter (preferably the phaseolin promoter (US 5504200), the USP (Baumlein, H. et al. Mol. Gen. Genet. (1991) 225 (3), 459 - 467) or LEB4 promoter (Fiedler and Conrad, 1995)), the LEB4 signal peptide, the gene to be expressed and an ER retention signal. The construction of such a cassette is shown in the form of a diagram in Figure 4 by way of example.

An expression cassette according to the invention is prepared by fusing a suitable promoter with a suitable HPPD DNA sequence and preferably a DNA which is inserted between promoter and HPPD DNA sequence and which encodes a chloroplast-specific transit peptide, and a polyadenylation signal, using customary recombination and cloning techniques as they are described, for example, in T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) and in T.J. Silhavy, M.L. Berman and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and in Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley-Interscience (1987).

Particularly preferred sequences are those which guarantee targeting into the apoplast, into plastids, into the vacuole, the mitochondrium, the endoplasmatic reticulum (ER), or, by means of the absence of suitable operative sequences, the remaining in the compartment of formation, the cytosol (Kermode, Crit. Rev. Plant Sci. 15, 4 (1996), 285 - 423). Localization in the ER has proved to be especially advantageous for the amount of protein

accumulation in transgenic plants (Schouten et al. , Plant Mol. Biol. 30 (1996), 781 - 792).

The invention also relates to expression cassettes whose DNA sequence encodes an HPPD fusion protein, a moiety of the fusion protein being a transit peptide which governs translocation of the polypeptide. Especially preferred are chloroplast-specific transit peptides which are cleaved enzymatically from the HPPD moiety after the HPPD gene product has been translocated into the chloroplasts. Particularly preferred is the transit peptide which is derived from plastid transketolase (TK) or a functional equivalent of this transit peptide (eg. the transit peptide of the small subunit of rubisco or of Ferredoxin NADP oxidoreductase).

The HPPD-encoding nucleotide sequence inserted can be prepared synthetically or obtained naturally or comprise a mixture of synthetic and natural DNA components. In general, there are prepared synthetic nucleotide sequences with codons which are preferred by plants. These codons which are preferred by plants can be determined from amongst codons with the highest protein frequency which are expressed in most interesting plant species. When preparing an expression cassette, various DNA fragments may be manipulated in order to obtain a nucleotide sequence which expediently reads in the correct direction and which is provided with a correct reading frame. To connect the DNA fragments to each other, adaptors or linkers may be joined onto the fragments.

The promoter and terminator regions according to the invention may advantageously be provided, in the direction of transcription, with a linker or polylinker which comprises one or more restriction sites for insertion of this sequence. As a rule, the linker has 1 to 10, in most cases 1 to 8, preferably 2 to 6, restriction sites. In general, the linker within the regulatory regions has a size of less than 100 bp, frequently less than 60 bp, but at least 5 bp. The promoter according to the invention may be both native, or homologous, but also foreign, or heterologous, to the host plant. The expression cassette according to the invention comprises, in the 5'-3' transcription direction, the promoter according to the invention, any desired DNA sequence and a region for transcriptional termination. Various termination regions can be exchanged for each other as desired.

It is furthermore possible to employ manipulations which provide suitable restriction sites or which remove excess DNA or restriction sites. Where insertions, deletions or substitutions,

eg. transitions and transversions, are suitable, it is possible to use *in vitro* mutagenesis, primer repair, restriction or ligation. In the case of suitable manipulations, eg. restriction, chewing back or filling up overlaps for blunt ends, complementary ends of the fragments may be provided for ligation.

What may be of importance for the success according to the invention is, inter alia, attaching the specific ER retention signal SEKDEL (Schouten, A. et al. Plant Mol. Biol. 30 (1996), 781 - 792), which results in a three to four times higher than average expression level. Other retention signals which occur naturally in plant and animal proteins which are localized in the ER may also be used for constructing the cassette.

Preferred polyadenylation signals are plant polyadenylation signals, preferably those which correspond essentially to T-DNA polyadenylation signals from *Agrobacterium tumefaciens*, in particular of gene 3 of the T-DNA (octopine synthase) of the Ti plasmid pTiACH5 (Gielen et al., EMBO J. 3 (1984) 835 et seq.), or functional equivalents.

An expression cassette according to the invention may comprise, for example, a constitutive promoter (preferably the CaMV 35S promoter), the LeB4 signal peptide, the gene to be expressed and the ER retention signal. The preferred ER retention signal used is the amino acid sequence KDEL (lysine, aspartic acid, glutamic acid, leucine).

The fused expression cassette which encodes an HPPD gene is preferably cloned into a vector, for example pBin19, which is suitable for transforming *Agrobacterium tumefaciens*. Agrobacteria which are transformed with such a vector can then be used in the known manner for transforming plants, in particular crop plants, eg. tobacco plants, for example by immersing scarified leaves or leaf sections in an agrobacteria solution and subsequently growing them in suitable media. The transformation of plants by means of agrobacteria is known, inter alia, from F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, Eds. S.D. Kung and R. Wu, Academic Press, 1993, pp. 15 - 38. The transformed cells of the scarified leaves or leaf sections can be used for regenerating, in the known manner, transgenic plants which contain a gene for expression of an HPPD gene integrated into the expression cassette according to the invention.

To transform a host plant with an HPPD-encoding DNA, an expression cassette according to the invention is incorporated into a recombinant vector in the form of an insertion, and the vector DNA of this recombinant vector additionally comprises functional regulation signals, for example sequences for replication or integration. Suitable vectors are described, inter alia, in "Methods in Plant Molecular Biology and Biotechnology" (CRC Press), Chapter. 6/7, pp. 71 - 119 (1993).

Using the above-cited recombination and cloning techniques, the expression cassettes according to the invention can be cloned into suitable vectors which allow their multiplication, for example in *E. coli*. Suitable cloning vectors are, inter alia, pBR332, pUC series, M13mp series and pACYC184. Especially suitable are binary vectors which are capable of replicating not only in *E. coli*, but also in agrobacteria.

The invention furthermore relates to the use of an expression cassette according to the invention for transforming plants, plant cells, plant tissues or plant organs. The preferred purpose of the use is to raise the vitamin E content of the plant.

Depending on the choice of the promoter, expression may take place specifically in the leaves, in the seeds or in other plant organs. The present invention also relates to such transgenic plants, their propagation material and their plant cells, plant tissues or plant organs.

In addition, the expression cassette according to the invention may also be employed for transforming bacteria, cyanobacteria, yeasts, filamentous fungi and algae with the purpose of raising the vitamin E production.

The transfer of foreign genes into the genome of a plant is termed transformation. This process exploits the previously described methods of transforming and regenerating plants from plant tissues or plant cells to obtain transient or stable transformation. Suitable methods are protoplast transformation by polyethylene-glycol induced DNA uptake, the ballistic method with the gene gun - the so-called particle bombardment method, electroporation, incubation of dry embryos in DNA-containing solution, microinjection and *Agrobacterium*-mediated gene transfer. The abovementioned methods are described, for example, in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, Eds. S.D. Kung and R. Wu, Academic Press (1993) 128 - 143, and in Potrykus Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991) 205 - 225). The

construct to be expressed is preferably cloned into a vector which is suitable for transforming *Agrobacterium tumefaciens*, for example pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984) 8711).

Agrobacteria transformed with an expression cassette according to the invention can also be used, in a known manner, for transforming of plants, in particular crop plants such as cereals, maize, oats, soya, rice, cotton, sugar beet, canola, sunflowers, flax, hemp, potatoes, tobacco, tomatoes, oilseed rape, alfalfa, lettuce and the various tree, nut and grapevine species, for example by immersing scarified leaves or leaf sections in an agrobacteria solution and subsequently growing them in suitable media.

Functionally equivalent sequences which encode an HPPD gene are, in accordance with the invention, those sequences which still have the desired functions despite a different nucleotide sequence. Thus, functional equivalents embrace naturally occurring variants of the sequences described herein and also artificial nucleotide sequences, eg. artificial nucleotide sequences which have been obtained by chemical synthesis and which are adapted to the codon usage of a plant.

A functional equivalent is also to be understood as meaning, in particular, natural or artificial mutations of an originally isolated HPPD-encoding sequence which continues to show the desired function. Mutations encompass substitutions, additions, deletions, inversions or insertions of one or more nucleotide residues. Thus, the present invention also encompasses those nucleotide sequences which are obtained by modifying the present nucleotide sequence. The purpose of such a modification may be, for example, a further limitation of the encoding sequence contained therein, or else, for example, the insertion of further cleavage sites for restriction enzymes.

Functional equivalents are also those variants whose function is less or more pronounced in comparison with the starting gene or gene fragment.

Also suitable are artificial DNA sequences as long as they, as described above, mediate the desired characteristic of raising the vitamin E content in the plant by overexpressing the HPPD gene in crop plants. Such artificial DNA sequences can be determined for example by back-translation of proteins constructed with the aid of molecular modeling and which have HPPD activity, or by *in vitro* selection. Especially suitable are encoding DNA sequences which were obtained by back-translating a

polypeptide sequence in accordance with the codon usage specific to the host plant. The specific codon usage can be determined readily by an expert familiar with plant genetic methods using computer evaluations of other, known genes of the plant to be transformed.

Further suitable equivalent nucleic acid sequences according to the invention which must be mentioned are sequences which encode fusion proteins, a component of the fusion protein being a plant HPPD polypeptide or a functionally equivalent moiety thereof. The second moiety of the fusion protein can be, for example, a further polypeptide with enzymatic activity or an antigenic polypeptide sequence with the aid of which the detection of HPPD expression is possible (eg. myc-tag or his-tag). However, this is preferably a regulatory protein sequence, eg. a signal or transit peptide, which leads the HPPD protein to the desired site of action.

However, the invention also relates to the expression products and fusion products, of a transit peptide and a polypeptide with HPPD activity, which have been produced in accordance with the invention.

Raising the vitamin E content means, for the purposes of the present invention, the artificially acquired ability of an elevated vitamin E biosynthesis performance by means of functional overexpression of the HPPD gene in the plant in contrast to the non-genetically-engineered plant for the duration of at least one plant generation.

The vitamin E biosynthesis site is generally the leaf tissue, so that leaf-specific expression of the HPPD gene is expedient. However, it will be understood readily that vitamin E biosynthesis is not necessarily restricted to the leaf tissue, but may also take place tissue-specifically in all other organs of the plant, for example in fatty seeds.

In addition, constitutive expression of the exogenous HPPD gene is advantageous. On the other hand, inducible expression may also appear desirable.

The efficacy of expression of the transgenically expressed HPPD gene can be determined for example *in vitro* by shoot meristem propagation. In addition, changes in the nature and level of HPPD gene expression, and its effect on the vitamin E biosynthesis

performance on test plants, can be tested in greenhouse experiments.

The invention furthermore relates to transgenic plants transformed with an expression cassette according to the invention, and to transgenic cells, tissues, organs and propagation material of such plants. Especially preferred in this context are transgenic crop plants, eg. barley, wheat, rye, maize, oats, soya, rice, cotton, sugar beet, canola, sunflowers, flax, hemp, potatoes, tobacco, tomatoes, oilseed rape, alfalfa, lettuce and the various tree, nut and grapevine species.

Plants for the purposes of the invention are mono- and dicotyledonous plants or algae.

As already mentioned, HPPD is a suitable target for sulcotrione-type herbicides. To allow even more efficient HPPD inhibitors, it is necessary to provide suitable test systems with which inhibitor/enzyme binding studies can be carried out. To this end, for example, the complete barley HPPD cDNA sequence is cloned into an expression vector (pQE, Qiagen) and overexpressed in *E. coli*.

The HPPD protein expressed with the aid of the expression cassette according to the invention is particularly suitable for finding HPPD-specific inhibitors.

To this end, the HPPD can be employed, for example, in an enzyme assay in which the HPPD activity is determined in the presence and absence of the active substance to be tested. A comparison of the two activity determinations allows qualitative and quantitative findings on the inhibitory behavior of the active substance to be tested to be obtained.

The test system according to the invention allows a large number of chemical compounds to be screened rapidly and simply for herbicidal properties. The method allows the targeted and reproducible selection, amongst a large number of substances, of those with great potency in order to subject these substances subsequently to further in-depth tests with which the expert is familiar.

The invention furthermore relates to herbicides which can be identified with the above-described test system.

Overexpression in a plant of the gene sequence Seq ID NO: 1, which encodes an HPPD, results in an elevated resistance to HPPD inhibitors. The invention also relates to the transgenic plants thus generated.

The invention furthermore relates to:

- A method of transforming a plant, which comprises introducing an expression cassette according to the invention into a plant cell, into callus tissue, into an entire plant or into plant protoplasts.

- The use of a plant for generating plant HPPD.

- The use of the expression cassette according to the invention for generating plants with elevated resistance to HPPD inhibitors by means of higher expression of a DNA sequence according to the invention.

- The use of the expression cassette according to the invention for generating plants with an elevated vitamin E content by means of expressing, in plants, a DNA sequence according to the invention.

- The use of the expression cassette according to the invention for generating a test system for identifying HPPD inhibitors.

The invention is illustrated by the examples which follow, but not limited thereto:

General cloning methods

The cloning steps carried out within the scope of the present invention, eg. restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids onto nitrocellulose and nylon membranes, linking DNA fragments, transformation of E. coli cells, growing bacteria, multiplying phages and sequence analysis of recombinant DNA, were carried out as described by Sambrook et al. (1989) Cold Spring Harbor Laboratory Press; ISBN 0-87969-309-6).

The bacterial strains used hereinbelow (*E. coli*, XL-I Blue) were obtained from Stratagene and, in the case of NP66, Pharmacia. The agrobacterial strain used for the transformation of plants (*Agrobacterium tumefaciens*, C58C1 with plasmid pGV2260 or pGV3850kann) was described by Deblaere et al. in (Nucl. Acids Res. 13 (1985) 4777). Alternatively, the agrobacterial strain LBA4404 (Clontech) or other suitable strains may also be employed. Vectors which can be used for cloning are the vectors pUC19 (Yanish-Perron, Gene 33 (1985), 103 - 119) pBluescript SK- (Stratagene), pGEM-T (Promega), pZerO (Invitrogen), pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984), 8711 - 8720) and pBinAR (Höfgen and Willmitzer, Plant Science 66 (1990) 221 - 230).

Sequence analysis of recombinant DNA

Recombinant DNA molecules were sequenced using a laser fluorescence DNA sequencer by Licor (available from MWG Biotech, Ebersbach) following the method of Sanger (Sanger et al., Proc. Natl. Acad. Sci. USA 74 (1977), 5463 - 5467).

Generation of plant expression cassettes

Into plasmid pBin19 (Bevan et al., Nucl. Acids Res. (1984) 12, 8711) there was inserted a 35S CaMV promoter in the form of an EcoRI-KpnI fragment corresponding to nucleotides 6909 - 7437 of cauliflower mosaic virus (Franck et al. Cell 21 (1980) 285). The polyadenylation signal of gene 3 of the T-DNA of the Ti plasmid pTiACH5 (Gielen et al., EMBO J. 3 (1984) 835), nucleotides 11749 - 11939, was isolated as a PvuII-HindIII fragment and, after addition of SphI linkers, cloned into the PvuII cleavage site between the SpHI-HindIII cleavage site of the vector pBmAR-Hyg. This gave the plasmid pBinAR (Höfgen and Willmitzer, Plant Science 66 (1990) 221 - 230).

Use Examples

Example 1

Isolation of HPPD-specific cDNA sequences

The composition of the mRNA population from primary leaves of nine-day-old barley plants which had been grown in an L/D photoperiod (16 hours light/8 hours dark) was compared with that of primary leaves of 11-day-old barley plants in which, after a raising period of nine days, senescence was subsequently induced by a two-day dark treatment (Humbeck and Krupinska, J. Photochem. Photobiol. 36 (1996), 321 - 326) with the aid of the DDRT-PCR method published by Liang and Pardee (Science (1992) 257, 967 - 972). In each case 0.2 µg of the total RNA was converted into cDNA using the enzyme "Superskript RT" (Gibco BRL, Eggenstein). In addition to the RNA, the reaction batches (20 µl) also contained 20 µM dNTPs, 10 µM DTT, 1xRT buffer and in each case 1 µM (dT)12VN primer. The anchor "primers" required for these reactions were synthesized on the basis of the data of Liang and Pardee:

1. 5'-TTTTTTTTTTTTAG-3'
2. 5'-TTTTTTTTTTTTCA-3'
3. 5'-TTTTTTTTTTTTAC-3'
4. 5'-TTTTTTTTTTTTGT-3'

After the cDNAs were synthesized, amplification of the relevant sequences was effected in each case in ten batches, which differ by the use of the random "primers" given hereinbelow:

1.	5'-TACAACGAGG-3'	2.	5'-GGAACCAATC-3'
3.	5'-AAACTCCGTC-3'	4.	5'-TGGTAAAGGG-3'
5.	5'-CTGCTTGATG-3'	6.	5'-GTTTTCGCAG-3'
7.	5'-GATCTCAGAC-3'	8.	5'-GATCTAACCG-3'
9.	5'-GATCATGGTC-3'	10.	5'-GATCTAAGGC-3'

In a volume of in each case 20 µl, the PCR reaction batches contained 1xPCR buffer, 2 µM dNTPs, 2.5 µCi (α ^{33}P)-dATP, 1 µM $(dT)_{12}VN$-"primer", 1/10 vol. RT mix (Sambrook et al. Molecular Cloning - A Laboratory Manual, 1989), 1 U Taq DNA polymerase (Boehringer, Mannheim) and 1 µM 10-mer random "primers". The PCR-reactions proceeded in a Uno block (Biometra) following the program below:

1. 94°C 2 min
2. 94°C 30 s

3. 40°C 2 min
4. 72°C 30 s
5. 72°C 5 min
6. 4°C storage until further processing

Steps 2, 3 and 4 were carried out 40 times in succession. This gave approximately 100 cDNA bands per reaction and "primer" combination.

In contrast to the protocol of Liang and Pardee, the amplified cDNA fragments were separated in non-denaturing polyacrylamide gels of the following composition: 6% (w/v) acrylamide (Long Ranger, AT Biochem), 1.2 x TBE buffer, 0.005% (v/v) TEMED and 0.005% (w/v) APS (Bauer et al, Nucl. Ac. Res. (1993) 21, 4272 - 4280).

In each case 3.5 µl of each PCR batch were treated with 2 µl of loading buffer (dye II, Sambrook et al., 1989) and then loaded onto the gel. To determine the reproducibility of the cDNA band patterns (Fig. 5), in each case two independent RNA preparations (9 and 9', 11 and 11') were prepared from the barley primary leaves harvested on days 9 and 11 and used in parallel in the analysis below. What is shown is the result of two different primer combinations (A and B); by way of example, two differences in the band pattern between the sample of days 9 and 11 were emphasized by arrows. Only those bands which occurred equally in the two samples from senescing plants and which did not occur in the two comparison samples were taken into consideration when analyzing the gels at a later point in time. Electrophoresis was carried out over a period of 2.5 hours at 40 watt (0.8 w/cm^3) in 1 x TBE buffer. After separation of the cDNA fragments had occurred, the gel was transferred onto filter paper (Schleicher & Schüll, Dassel). After the gel had been dried at 50°C, an X-ray film was placed on top of it. cDNA bands which were only found in the case of samples 11 and 11' in the autoradiograph were excised from the dry gel using a surgical blade, and the DNA was eluted by boiling in 100 µl 1 x TE buffer. The ethanol-precipitated DNA was resuspended in 10 µl of water for further tests. After reamplification with the "primers" previously used for this batch, the DNA was cloned and sequenced and also employed as a probe for Northern blot hybridizations.

To test if the relevant cDNA fragment actually represents a senescence-specifically occurring transcript, hybridizations were carried out with RNA from leaves of various developmental stages:

A. 1. RNA from primary leaves from plants raised for 9 days in an L/D photoperiod

A. 3. RNA from primary leaves from 10-day-old plants raised without a light phase on day 10

A. 4 RNA from primary leaves from 11-day-old plants which lacked a light phase on days 10 and 11

A. 5 RNA from primary leaves from 12-day-old plants which underwent a further light phase after 2 days in the dark

The samples for RNA analysis were harvested in each case in the middle of the original night phase.

B. RNA from flag leaves which had been collected in the field at seven different points in time (Fig. 6). The leaves were fully grown on 29 May and showed less than 10% of the original chlorophyll content on 21 June. The beginning of the senescence processes is shown in Figure 6 by an arrow (ie. 17 days after reaching the full length on 15 June). The beginning of senescence was defined as the day on which photosystem II efficacy dropped (Humbeck et al., Plant Cell Environment (1996) 19: 337 - 344).

To hybridize a filter with the above-described RNA samples, a specific probe for the *rbcS* gene, which encodes the small sub-unit of ribulose-1,5-bisphosphate carboxylase, was also employed in addition to the HPPD probe, for comparison reasons. Figure 6 shows hybridization of the "Northern blots" A and B with cDNA HvSD36 and with a probe which is specific for the rbcS gene. Filter A carries RNA from barley primary leaves after a raising period of 9 days in an L/D photoperiod (9), after subsequent incubation in the dark for one and two days, respectively (10, 11) and after subsequent return to light conditions for one day (12). Filter B contains RNA from flag leaves which had been harvested in the field in the period from 29.05. to 21.06.1992. The arrow indicates the beginning of senescence on 15.06. As can been seen from Figure 6, the amount of *rbcS*-specific mRNA is high when the amount of HPPD-specific mRNA is relatively low. In primary leaves of nine-day-old plants, the HPPD-specific mRNA is not detectable prior to transfer into the dark and accumulates markedly during the dark phase. When the plants are returned to light conditions, the amount of this mRNA drops markedly. In the case of the flag leaves, small amounts of the HPPD-specific mRNA can already be detected in fully-grown, non-senescent leaves. As

early as 4 days prior to the actual beginning of senescence, expression levels are higher. The highest amount of this mRNA can be found in senescent leaves. A size comparison with known RNA species showed that the transcript detected with the cDNA probe HvSD36 (s: senescence; d: dark, fragment number 36 in the DDRT gel) has a length of approx. 1.6 kb.

By means of DDRT PCR, three cDNA fragments were obtained independently of each other which showed this expression pattern and which, on the basis of sequence analysis, actually represent the same transcript. The longest fragment had a size of 230 bp. The 230 bp long PCR product was finally cloned into the SmaI cleavage site of vector pUC18 using the "Sure Clone™ ligation kit" (Pharmacia, Freiburg) following the manufacturer's instructions. The recombinant plasmid was transformed into competent cells of E. coli strain DH5 α. Since, for methodology reasons, the fragment represents the 3' end of the relevant transcript, the sequence information was first insufficient to identify an unambiguous homology with a sequence in the databases. To isolate a longer corresponding cDNA, a lambda ZAPII library (Stratagene, Heidelberg) of RNA of senescent flag leaves was screened using the 230 bp long fragment as the probe. For this step, the probe was labeled with Dig-dUTP following the instructions of the "DNA Labeling and Detection Kit" (Boehringer, Mannheim). The library was examined following the protocol of the "ZAP-cDNA Synthesis Kit" (Stratagene, Heidelberg).

In the case of the probe described herein, 150,000 pfu were examined. Of these, 39 phage plaques gave a positive signal. Of these, further work was carried out on 12 phage populations. Following phage preparation, the fragments inserted were enriched via PCR and separated by electrophoresis. Southern blot hybridization with the HvSD36 probe allowed those phage populations which had the largest "inserts" with positive signal to be selected amongst the 12 phage populations thus treated. After replating, the phages were subjected to a further hybridization step. Single phage plaques were excised and, after elution, subjected to an *in vivo* excision using a helper phage and following the protocol from Stratagene (Exassist™ Interference-Resistant Helper Phage with SOLR ™ Strain). The so-called "phagemids" obtained from this treatment contain the cDNA cloned in pBLueskript (SK-).

Following a subsequent plasmid preparation, the relevant "insert" was excised from the Bluescript plasmid using EcoRI. The cDNA clone obtained in the case of HvSD36 contains an "insert" with a length of approx. 800 bp. Complete sequencing of the cDNA was

carried out using the "SequiTherm Excel Long-Read DNA-Sequenzierungs-Kit" (Epicentre Technologies, Biozym Diagnostic, Oldendorf) using IRD41-labeled universal "primers" which bind to sequence regions in the Bluescript vector. Detection of the DNA fragments was effected via the infrared laser of the automatic sequencer 4000L by Licor. After sequencing, an exactly 759 bp long sequence was present whose sides are flanked by an in each case 14 bp long adaptor sequence. These adaptor sequences were used for ligating the cDNA fragments with the arms of phage lambda ZAPII (Stratagene, Heidelberg) when generating the c-DNA library.

Amongst the sequences in the databases, the protein sequence HvSD36, which has a total of over 180 amino acids, has a homology of 41% with the sequence of human HPPD which is the highest. Taking into consideration the length of the transcript detected in the "Northern blot" (approx. 1600 nt), it can be assumed that 850-900 bp are still missing from the cDNA.

To complete the cDNA, a further cDNA library was investigated. mRNA was isolated from the basal meristematic zone of 5-day-old barley seedlings with the aid of "Dynabeads" (Dynal, Hamburg) and transcribed into cDNA using the "Time Saver cDNA SyntheseKit" (Pharmacia, Freiburg). This was followed by ligation of EcoRI/NotI adaptors (Pharmacia, Freiburg) to the cDNA with subsequent ligation into the lambda ExCell vector (Pharmacia, Freiburg). Finally, the recombinant phage DNA was packaged into phage proteins with the aid of "Gigapack II Gold Set" (Stratagene, Heidelberg). Using the 759 bp long probe HvSD36, 400,000 pfu were screened, and 5 phages were detected by the probe. Excision of the "phagemids" from the phage was effected *in vivo* with the aid of bacterial strain NP66 following the instructions of Pharmacia (Freiburg). The recombinant pExCell plasmids were isolated from the individual bacterial colonies and transferred into bacterial strain D115 α for propagation.

The longest cDNA clone HvSD36 isolated in this manner has a length of 1565 bp and was sequenced completely (see sequence protocol).

Example 2

Characterization of the genomic sequence

To identify a genomic clone which contains the gene sequence of HPPD, a lambda FIXII library of barley was obtained from Stratagene (Heidelberg). The library was prepared using DNA from

etiolated leaves of winter barley cv. Igri. The DNA was partially digested with Sau3AI. Prior to cloning into the XhoI cleavage site of the vector, the fragment ends and the phage arms were filled up with nucleotides. Screening of the library with 200,000 pfu in the first round only gave one clone which hybridized with cDNA HvSD36. After subjecting this recombinant phage to restriction digestion with PstI and SacI, fragments 5400, 3800 and 1800 bp in length were subsequently isolated which can be detected with the HvSD36 probe when carrying out a "Southern" blot hybridization. These sub-fragments exist in cloned form in the Bluescript vector.

The library was screened following the protocol given for the HybondN membrane. Labeling of the probe for screening the library and for the "Southern" blot hybridizations was effected via "random priming" with ^{32}P-dATP using the Klenow enzyme (Sambrook et al., (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, New York).

A genomic "Southern blot" was carried out with total DNA from barley (Carina) (Fig. 7). In each case 15 µg of DNA were digested with BamHI (B), EcoRI (E), HindIII (H) or XBAI (X) and separated in a 0.75% agarose gel. After transfer to a Hybond N+ membrane (Amersham, Braunschweig), hybridization was effected with the incomplete, 759 bp long cDNA probe from HvSD36 following instructions of the membrane manufacturer. The following fragments were detected:

BamHI:	6.0, 3.9 and 3.0 kbp
EcoRI:	>10 kbp
HindIII:	8.3, 2.6, 1.1 and 1.0 kbp
XbaI:	9.0, 5.2 and 4.2 kbp

The fragment lengths were estimated by comparison with a DNA size standard (Kb-Leiter, GibcoBRL, Eggenstein).

Example 3

Homology comparison of the HvSD36 protein sequence

A comparison of the HvSD36 protein sequence with protein sequences in the database revealed homologies to the following protein sequences known to date:

```
                        10         20         30         40         50
HPPD_Hv       .......... .......... .......... ........MP PTPTTPAATG
HPPD_Ath      .......... .......... .......... ...MGHQNAA VSENQNHDDG
HPPD_HUMAN    .......... .......... .......... .......... ..........
HPPD_RAT      .......... .......... .......... .......... ..........
HPPD_PIG      .......... .......... .......... .......... ..........
HPPD_MOUSE    .......... .......... .......... .......... ..........
HPPD_PSESP    .......... .......... .......... .......... ..........
MELA_SHECO    .......... .......... .......... .......... ..........
PEA3_MOUSE    MTKSSNHNCL LRPENKPGLW GPGAQAASLR PSPATLVVSS PGHAEHPPAA

                        60         70         80         90        100
HPPD_Hv       AAAAVTPEHA RPHRMVRFNP RSDRFHTLSF HHVEFWCADA ASAAGRFAFA
HPPD_Ath      AASSPGFKLV GFSKFVRKNP KSDKFKVKRF HHIEFWCGDA TNVARRFSWG
HPPD_HUMAN              M TTYSDKGAKP ERGRFLH--F HSVTFWVGNA KQAASFYCSK
HPPD_RAT                  YWDKGPKP ERGRFLH--F HSVTFWVGNA KQAASFYCNK
HPPD_PIG                M TSYSDKGEKP ERGRFLH--F HSVTFWVGNA KQAASYYCSK
HPPD_MOUSE              M TTYNNKGPKP ERGRFLH--F HSVTFWVGNA KQAASFYCNK
HPPD_PSESP                             ADLYENP MGLMGFEFIE LASPTPNTLE
MELA_SHECO                             MASEQNP LGLLGIEFTE FATPDLDFMH
PEA3_MOUSE    PAQTPGPQVS ASARGPGPVA GGSGRMERRM KGGYL---DQ RVPYTFCSKS

                       110        120        130        140        150
HPPD_Hv       LGAPLAARSD LSTGNSAHAS QLLRSGSLAF LFT--APYAN G-CDAA----
HPPD_Ath      LGMRFSAKSD LSTGNMVHAS YLLTSGDLRF LFT--APYSP S-LSAGEIKP
HPPD_HUMAN    MGFEPLAYRG LETGSREVVS HVIKQGKIVF VLS--SA--- -------LNP
HPPD_RAT      MGFEPLAYKG LETGSREVVS HVIKQGKIVF VLC--SA--- -------LNP
HPPD_PIG      IGFEPLAYKG LETGSREVVS HVVKQDKIVF VFS--SA--- -------LNP
HPPD_MOUSE    MGFEPLAYRG LETGSREVVS HVIKRGKIVF VLC--SA--- -------LNP
HPPD_PSESP    PIFEIMGFTK VATHRSKDV- HLYRQGAINL ILN--NE--- ----------
MELA_SHECO    KVFIDFGFSK LKKHKQKDI- VYYKQNDINF LLN--NE--- ----------
PEA3_MOUSE    PGNGSLGEAL MVPQGKLMDP GSLPPSDSED LFQDLSHFQE TWLAEAQVPD

                       160        170        180        190        200
HPPD_Hv       --TASLPSFS ADAARRFSAD HGIAVRSVAL RVADAAEAFR ASRRRGARPA
HPPD_Ath      TTTASIPSFD HGSCRSFFSS HGLGVRAVAI EVEDAESAFS ISVANGAIPS
HPPD_HUMAN    --------WN KEMGDHL-VK HGDGVKDIAF EVEDCDYIVQ KARERGAKIM
HPPD_RAT      --------WN KEMGDHL-VK HGDGVKDIAF EVEDCEHIVQ KARERGAKIV
HPPD_PIG      --------WN KEMGDHL-VK HGDGVKDIAF EVEDCDYIVQ KARERGAIIV
HPPD_MOUSE    --------WN KEMGDHL-VK HGDGVKDIAF EVEDCDHIVQ KARERGAKIV
HPPD_PSESP    ---------P HSVASYFAAE HGPSVCGMAF RVKDSQKAYK RALELGAQPI
MELA_SHECO    ---------K QGFSAQFAKT HGPAISSMGW RVEDANFAFE GAVARGAKPA
PEA3_MOUSE    SDEQFVPDFH ---SENLAFH SPTTRIKKEP QSPRTDPALS CSRKPPLPYH

                       210        220        230        240        250
HPPD_Hv       FAPV------ -----DLGRG FAFAEVELYG --DVVLRFVS HP--DG--TD
HPPD_Ath      SPPI------ -----VLNEA VTIAEVKLYG --DVVLRYVS YKAEDT--EK
HPPD_HUMAN    REP------- -WVEQDKFGK VKFAVLQTYG --DTTHTLVE KMN-----YI
HPPD_RAT      REP------- -WVEEDKFGK VKFAVLQTYG --DTTHTLVE KIN-----YT
HPPD_PIG      REEVC-CAAD VRGHHTPLDR AR----QVWE --GT---LVE KMT-----FC
HPPD_MOUSE    REP------- -WVEQDKFGK VKFAVLQTYG --DTTHTLVE KIN-----YT
HPPD_PSESP    HI-------- ----ETGPME LNLPAIKGIG --GAPLYLID RFGEGSSIYD
MELA_SHECO    AD-------- ----EV--KD LPYPAIYGIG --DSLIYFID TFGDDNNIYT
PEA3_MOUSE    HGEQCLYSRQ IAIKSPAPGA PGQSPLQPFS RAEQQQSLLR ASSSSQSHPG
```

```
                    260        270        280        290        300
HPPD_Hv      VPFLPGFEGV TNPDA----- VDYGLTRFDH VVGNVP--EL -APAAAYIAG
HPPD_Ath     SEFLPGFERV EDASSF---P LDYGIRRLDH AVGNVP--EL -GPALTYVAG
HPPD_HUMAN   GQFLPGYEAP AFMDPLLPKL PKCSLEMIDH IVGNQPDQEM -VSASEW---
HPPD_RAT     GRFLPGFEAP TYKDTLLPKL PSCNLEIIDH IVGNQPDQEM -ESASEW---
HPPD_PIG     LDSRPQPSQT LLHRLLLSKL PKCGLEIIDH IVGNQPDQEM -ESASQW---
HPPD_MOUSE   GRFLPGFEAP TYKDTLLPKL PRCNLEIIDH IVGNQPDQEM -QSASEW---
HPPD_PSESP   IDFV--FLEG VDRHPVGA-- ---GLKIIDH LTHNVYRGRM -A---YWANF
MELA_SHECO   SDF-----EA LDEPIITQ-- -EKGFIEVDH LTNNVHKGTM -E---YWSNF
PEA3_MOUSE   HGYLGEHSSV FQQPVDMCHS FTSPQGGGRE PLPAPYQHQL SEPCPPYPQQ

                    310        320        330        340        350
HPPD_Hv      FT---GFHEF AEFTAEDVGT TESGLNSVVL ANNSEGVLLP LNEPVHGTKR
HPPD_Ath     FT---GFHQF AEFTADDVGT AESGLNSAVL ASNDEMVLLP INEPVHGTKR
HPPD_HUMAN   YLKNLQFHRF WSVDDTQVHT EYSSLRSIVV ANYEESIKMP INEPAPG-KK
HPPD_RAT     YLKNLQFHRF WSVDDTQVHT EYSSLRSIVV ANYEESIKMP INEPAPG-RK
HPPD_PIG     YMRNLQFHRF WSVDDTQIHT EYSALRSVVM ANYEESIKMP INEPAPG-KK
HPPD_MOUSE   YLKNLQFHRF WSVDDTQVHT EYSSLRSIVV TNYEESIKMP INEPAPG-RK
HPPD_PSESP   YEKLFNFREI RYF---DIKG EYTGLTSKAM TAPDGMIRIP LNE--ESSKG
MELA_SHECO   YKDIFGFTEV RYF---DIKG SQTALISYAL RSPDGSFCIP INE--GKGDD
PEA3_MOUSE   NFKQ-EYHDP LYEQAGQPAS SQGGVSGHRY PGAGVVIKQE RTDFAYDSDV

                    360        370        380        390        400
HPPD_Hv      RSQIQTFLEH HGGPGVQH-I AVASSDVLRT LRKMRARSAM GGFDFLPPPL
HPPD_Ath     KSQIQTYLEH NEGAGLQH-L ALMSEDIFRT LREMRKRSSI GGFDFMPSPP
HPPD_HUMAN   KSQIQEYVDY NGGAGVQH-I ALKTEDIITA IRHLRER--- -GLEFLSVP-
HPPD_RAT     KSQIQEYVDY NGGAGVQH-I ALRTEDIITT IRHLRER--- -GMEFLAVP-
HPPD_PIG     KSQIQEYVDY NGGAGVQH-I ALKTEDIITA IRSLRER--- -GVEFLAVP-
HPPD_MOUSE   KSQIQEYVDY NGGAGVQH-I ALKTEDIITA IRHLRER--- -GTEFLAAP-
HPPD_PSESP   AGQIEEFLMQ FNGEGIQH-V AFLSDDLIKT WDHLKSI--- -GMRFMTAPP
MELA_SHECO   RNQIDEYLKE YDGPGVQH-L AFRSRDIVAS LDAMEGS--- -SIQTLDIIP
PEA3_MOUSE   PGCASMYLHP EGFSGPSPGD GVMGYGYEKS LRPFPDDVCI VPKKFEGDIK

                    410        420        430        440        450
HPPD_Hv      PKYYEGVRRL AGD--VLSEA QIKECQELGV LVDRDDQG-- --VLL-----
HPPD_Ath     PTYYQNLKKR VGD--VLSDD QIKECEELGI LVDRDDQG-- --TLL-----
HPPD_HUMAN   STYYKQLREK LKTAKIKVKE NIDALEELKI LVDYDEKG-- --YLL-----
HPPD_RAT     SSYYRLLREN LKTSKIQVKE NMDVLEELKI LVDYDEKG-- --YLL-----
HPPD_PIG     FTYYKQLQEK LKSAKIRVKE SIDVLEELKI LVDYDEKG-- --YLL-----
HPPD_MOUSE   SSYYKLLREN LKSAKIQVKE SMDVLEELHI LVDYDEKG-- --YLL-----
HPPD_PSESP   DTYYEMLEGR LPN-----HGE PVGELQARGI LLDGSSESGD KRLLL-----
MELA_SHECO   E-YYDTIFEK LPQ-----VTE DRDRIKHHQI LVDGDEDG-- --YLL-----
PEA3_MOUSE   QEGIGAFREG PPYQR----- -RGALQLWQF LVALLDDPTN AHFIAWTGRG

                    460        470        480        490        500
HPPD_Hv      QIFTKPVGDR PTLFLEMIQR IGCMEKDERG EE-----YQKG GCGGFGKGNF
HPPD_Ath     QIFTKPLGDR PTIFIEIIQR VGCMMKDEEG KA-----YQSG GCGGFGKGNF
HPPD_HUMAN   QIFTKPVQDR PTLFLEVIQR HNHQ------ ---------- ---GFGAGNF
HPPD_RAT     QIFTKPMQDR PTLFLEVIQR HNHQ------ ---------- ---GFGAGNF
HPPD_PIG     QIFTKPMQDR PTVFLEVIQR NNHQ------ ---------- ---GFGAGNF
HPPD_MOUSE   QIFTKPMQDR PTLFLEVIQR HNHQ------ ---------- ---GFGAGNF
HPPD_PSESP   QIFSETLMGP --VFFEFIQR -----KGDD- ---------- ---GFGEGNF
MELA_SHECO   QIFTKNLFGP --IFIEIIQR -----KNNL- ---------- ---GFGEGNF
PEA3_MOUSE   MEFKLIEPEE VARLWGIQKN RPAMNYDKLS RSLRYYYEKG IMQKVAGERY

                    510        520        530        540        550
HPPD_Hv      --------SE LFK-SIE-DY --EKS--LEA KQSAAV-QGS
```

```
HPPD_Ath          --------SE LFK-SIE-EY --EKT--LEA KQLVG
HPPD_HUMAN        --------NS LFK-AFEEEQ --NLRGNLTN METNGVVPGM
HPPD_RAT          --------NS LFK-AFEEEQ --ALRG
HPPD_PIG          --------NS LFK-AFEEEQ --ELRGNLTD TDPNGVPFRL
HPPD_MOUSE        --------NS LFK-AFEEEQ --ALRGNLTD LEPNGVRSGM
HPPD_PSESP        --------KA LFE-SIERDQ --VRRGVLST -D
MELA_SHECO        --------KA LFE-SIERDQ --VRRGVL
PEA3_MOUSE        VYKFVCEPEA LFSLAFPDNQ RPALKAEFDR PVSEEDTVPL SHLDESPAYL

                                  560        570

HPPD_Hv
HPPD_Ath
HPPD_HUMAN
HPPD_RAT
HPPD_PIG
HPPD_MOUSE
HPPD_PSESP
MELA_SHECO
PEA3_MOUSE        PELTGPAPPF GHRGGYSY
```

Key: HPPD_Hv: *Hordeum vulgare* 4-hydroxyphenylpyruvate dioxygenase (HvSD36)
HPPD_Ath: *Arabidopsis thaliana* 4-hydroxyphenylpyruvate dioxygenase
HPPD_HUMAN: *H.sapiens* 4-hydroxyphenylpyruvate dioxygenase
HPPD_PIG: pig 4-hydroxyphenylpyruvate dioxygenase
HPPD_RAT: rat F alloantigen
HPPD_MOUSE: mouse 4-hydroxyphenylpyruvate dioxygenase
MELA_SHECO: *S. colwelliana* melA protein
HPPD_PSESP: *Pseudomonas* sp. (strain P.J.874) 4-hydroxyphenylpyruvate dioxygenase
PEA3_MOUSE: *Mus musculus* (mouse) PEA3 polypeptide

The greatest homology was with the Arabidopsis sequence, viz. 58% over the entire sequence (62% over 412 amino acids), followed by HPPD_RAT with 35% (over 365 amino acids), HPPD_HUMAN 34% (over 365 amino acids), HPPD_MOUSE 34% (over 371 amino acids).

Example 4

Raising barley (*Hordeum vulgare*)

Barley seedlings (*Hordeum vulgare* L. cv. Carina, Ackermann Saatzucht, Irbach, Germany) were raised over a period of 15 days under controlled conditions in a controlled-environment cabinet in so-called Mitscherlich pots in soil containing 4 g of Osmocote 5M (Urania, Hamburg, Germany) per liter. To ensure uniform growth, the seeds were germinated on moist filter paper in the dark for 2 days at 4°C and 1 day at 21°C, and only those seedlings

were planted which showed the same longitudinal growth of the primary root. After these seedlings had been transferred onto soil, they were covered with screened soil to a depth of 1.5 cm. Thereafter, the plants were incubated for 9 days at 16 hours light (120 $\mu m \cdot m^{-2} \cdot s^{-1}$) and 8 hours darkness in conjunction with a temperature shift (21°C during the day, 16°C during the night). After 9 days, the plants were kept for 2 days (days 10 and 11) in the dark at the abovementioned temperature in order to induce senescence.

Example 5
Raising tobacco

The tobacco plants were raised following the known method. The tobacco cultivar used is *Nicotiana tabacum* cv. Xanthi.

Example 6
Transformation of tobacco

The expression cassette according to the invention comprising the HPPD gene with Sequence 1 was cloned into vector pBinAR-Hyg (Fig. 4). Tobacco plants as described in Example 5 were subsequently transformed with this vector following the known method.

Example 7
Increasing the tocopherol biosynthesis in tobacco

The HPPD cDNA was provided with a CaMV 35S promoter and overexpressed in tobacco using the 35S promoter. In parallel, the seed-specific phaseolin gene promoter was used to increase the tocopherol content specifically in the tobacco seed. Tobacco plants which had been transformed with the relevant constructs were raised in the greenhouse. The α-tocopherol content of the total plant and of the seeds of the plant was subsequently determined. In all cases, the α-tocopherol concentration was increased in comparison with the untransformed plant.

SEQUENCE PROTOCOL

(1) GENERAL INFORMATION

(i) APPLICANT

(A) NAME: BASF AG

(B) STREET: Carl Bosch

(C) TOWN: Ludwigshafen

(D) FEDERAL COUNTRY: Germany

(F) POSTCODE: 67056

(G) TELEPHONE: 0621-60-52698

(ii) TITLE OF APPLICATION: HPPD sequence from barley

(iii) NUMBER OF SEQUENCES: 2

(iv) COMPUTER-READABLE FORM:

(A) RECORDING MEDIUM: floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn release #1.0, Version #1.25 (EPA)

(2) INFORMATION ON SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1565 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETIC: NO

(iii) ANTISENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: hppd from barley

(D) DEVELOPMENTAL STAGE: senescence

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: lambda FIXII library of barley

(B) CLONE: pHvSD36.seq

(ix) FEATURES:

(A) NAME/KEY: CDS

(B) POSITION: 9..1313

(x) PUBLICATION DETAILS:

(A) AUTHORS: Krupinska, Karin

(B) TITLE: Overexpression of HPPD

(C) JOURNAL: overexpression of HPPD

(G) DATE: 1998

(K) RELEVANT RESIDUES IN SEQ ID NO: 1 FROM 1 TO 1565

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CGCACACC ATG CCG CCC ACC CCC ACC ACC CCC GCG GCT ACC GGC GCC GCC      50
         Met Pro Pro Thr Pro Thr Thr Pro Ala Ala Thr Gly Ala Ala
          1               5                  10
```

```
GCC GCG GTG ACG CCG GAG CAC GCG CGA CCG CAC CGA ATG GTC CGC TTC          98
Ala Ala Val Thr Pro Glu His Ala Arg Pro His Arg Met Val Arg Phe
 15                  20                  25                  30

AAC CCG CGC AGC GAC CGC TTC CAC ACG CTC TCC TTC CAC CAC GTC GAG         146
Asn Pro Arg Ser Asp Arg Phe His Thr Leu Ser Phe His His Val Glu
                 35                  40                  45

TTC TGG TGC GCG GAC GCC GCC TCC GCC GCC GGC CGC TTC GCG TTC GCG         194
Phe Trp Cys Ala Asp Ala Ala Ser Ala Ala Gly Arg Phe Ala Phe Ala
             50                  55                  60

CTC GGC GCG CCG CTC GCC GCC AGG TCC GAC CTC TCC ACG GGG AAC TCC         242
Leu Gly Ala Pro Leu Ala Ala Arg Ser Asp Leu Ser Thr Gly Asn Ser
         65                  70                  75

GCG CAC GCC TCC CAG CTG CTC CGC TCG GGC TCC CTC GCC TTC CTC TTC         290
Ala His Ala Ser Gln Leu Leu Arg Ser Gly Ser Leu Ala Phe Leu Phe
     80                  85                  90

ACC GCG CCC TAC GCC AAC GGC TGC GAC GCC GCC ACC GCC TCC CTG CCC         338
Thr Ala Pro Tyr Ala Asn Gly Cys Asp Ala Ala Thr Ala Ser Leu Pro
 95                  100                 105                 110

TCC TTC TCC GCC GAC GCC GCG CGC CGG TTC TCC GCC GAC CAC GGG ATC         386
Ser Phe Ser Ala Asp Ala Ala Arg Arg Phe Ser Ala Asp His Gly Ile
                115                 120                 125

GCG GTG CGC TCC GTA GCG CTG CGC GTC GCA GAC GCC GCC GAG GCC TTC         434
Ala Val Arg Ser Val Ala Leu Arg Val Ala Asp Ala Ala Glu Ala Phe
            130                 135                 140

CGC GCC AGT CGT CGA CGG GGC GCG CGC CCG GCC TTC GCC CCC GTG GAC         482
Arg Ala Ser Arg Arg Arg Gly Ala Arg Pro Ala Phe Ala Pro Val Asp
        145                 150                 155

CTC GGC CGC GGC TTC GCG TTC GCG GAG GTC GAG CTC TAC GGC GAC GTC         530
Leu Gly Arg Gly Phe Ala Phe Ala Glu Val Glu Leu Tyr Gly Asp Val
    160                 165                 170

GTG CTC CGC TTC GTC AGC CAC CCG GAC GGC ACG GAC GTG CCC TTC TTG         578
Val Leu Arg Phe Val Ser His Pro Asp Gly Thr Asp Val Pro Phe Leu
175                 180                 185                 190

CCG GGG TTC GAG GGC GTA ACC AAC CCG GAC GCC GTG GAC TAC GGC CTG         626
Pro Gly Phe Glu Gly Val Thr Asn Pro Asp Ala Val Asp Tyr Gly Leu
                195                 200                 205

ACG CGG TTC GAC CAC GTC GTC GGC AAC GTC CCG GAG CTT GCC CCC GCC         674
Thr Arg Phe Asp His Val Val Gly Asn Val Pro Glu Leu Ala Pro Ala
            210                 215                 220
```

```
GCA GCC TAC ATC GCC GGG TTC ACG GGG TTC CAC GAG TTC GCC GAG TTC      722
Ala Ala Tyr Ile Ala Gly Phe Thr Gly Phe His Glu Phe Ala Glu Phe
        225                 230                 235

ACG GCG GAG GAC GTG GGC ACG ACC GAG AGC GGG CTC AAC TCG GTG GTG      770
Thr Ala Glu Asp Val Gly Thr Thr Glu Ser Gly Leu Asn Ser Val Val
    240                 245                 250

CTC GCC AAC AAC TCG GAG GGC GTG CTG CTG CCG CTC AAC GAG CCG GTG      818
Leu Ala Asn Asn Ser Glu Gly Val Leu Leu Pro Leu Asn Glu Pro Val
255                 260                 265                 270

CAC GGC ACC AAG CGC CGG AGC CAG ATA CAG ACG TTC CTG GAA CAC CAC      866
His Gly Thr Lys Arg Arg Ser Gln Ile Gln Thr Phe Leu Glu His His
                275                 280                 285

GGC GGC CCG GGC GTG CAG CAC ATC GCG GTG GCC AGC AGT GAC GTG CTC      914
Gly Gly Pro Gly Val Gln His Ile Ala Val Ala Ser Ser Asp Val Leu
            290                 295                 300

AGG ACG CTC AGG AAG ATG CGT GCG CGC TCC GCC ATG GGC GGC TTC GAC      962
Arg Thr Leu Arg Lys Met Arg Ala Arg Ser Ala Met Gly Gly Phe Asp
        305                 310                 315

TTC CTG CCA CCC CCG CTG CCG AAG TAC TAC GAA GGC GTG CGA CGC CTT     1010
Phe Leu Pro Pro Pro Leu Pro Lys Tyr Tyr Glu Gly Val Arg Arg Leu
    320                 325                 330

GCC GGG GAT GTC CTC TCG GAG GCG CAG ATC AAG GAA TGC CAG GAG CTG     1058
Ala Gly Asp Val Leu Ser Glu Ala Gln Ile Lys Glu Cys Gln Glu Leu
335                 340                 345                 350

GGT GTG CTC GTC GAT AGG GAC GAC CAA GGG GTG TTG CTC CAA ATC TTC     1106
Gly Val Leu Val Asp Arg Asp Asp Gln Gly Val Leu Leu Gln Ile Phe
                355                 360                 365

ACC AAG CCA GTA GGG GAC AGG CCG ACC TTG TTC CTG GAG ATG ATC CAG     1154
Thr Lys Pro Val Gly Asp Arg Pro Thr Leu Phe Leu Glu Met Ile Gln
            370                 375                 380

AGG ATC GGG TGC ATG GAG AAG GAC GAG AGA GGG GAA GAG TAC CAG AAG     1202
Arg Ile Gly Cys Met Glu Lys Asp Glu Arg Gly Glu Glu Tyr Gln Lys
        385                 390                 395

GGT GGC TGC GGC GGG TTC GGC AAA GGC AAC TTC TCC GAG CTG TTC AAG     1250
Gly Gly Cys Gly Gly Phe Gly Lys Gly Asn Phe Ser Glu Leu Phe Lys
    400                 405                 410

TCC ATT GAA GAT TAC GAG AAG TCC CTT GAA GCC AAG CAA TCT GCT GCA     1298
Ser Ile Glu Asp Tyr Glu Lys Ser Leu Glu Ala Lys Gln Ser Ala Ala
415                 420                 425                 430
```

GTT CAG GGA TCA TAGGATAGAA GCTGGTCCTT GTATCATGGT CTCATGGAGC 1350
Val Gln Gly Ser
435

AAAAGAAAAC AATGTTGTTT GTAATATGCG TCGCACAATT ATATCAATGT TATAATTGGT 1410

GAAGCTGAAG ACAGATGTAT CCTATGTATG ATGGGTGTAA TGGATGGTAG AGGGGCTCAC 1470

ACATGAAGAA AATGTAGCGT TGACATTGTT GTACAATCTT GCTTGCAAGT AAAATAAAGA 1530

ACAGATTTTG AGTTCTGCAA AAAAAAAAAA AAAAA 1565

(2) INFORMATION ON SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Pro Pro Thr Pro Thr Thr Pro Ala Ala Thr Gly Ala Ala Ala Ala
1 5 10 15

Val Thr Pro Glu His Ala Arg Pro His Arg Met Val Arg Phe Asn Pro
20 25 30

Arg Ser Asp Arg Phe His Thr Leu Ser Phe His His Val Glu Phe Trp
35 40 45

Cys Ala Asp Ala Ala Ser Ala Ala Gly Arg Phe Ala Phe Ala Leu Gly
50 55 60

Ala Pro Leu Ala Ala Arg Ser Asp Leu Ser Thr Gly Asn Ser Ala His
65 70 75 80

Ala Ser Gln Leu Leu Arg Ser Gly Ser Leu Ala Phe Leu Phe Thr Ala
85 90 95

Pro Tyr Ala Asn Gly Cys Asp Ala Ala Thr Ala Ser Leu Pro Ser Phe
100 105 110

Ser Ala Asp Ala Ala Arg Arg Phe Ser Ala Asp His Gly Ile Ala Val
115 120 125

Arg Ser Val Ala Leu Arg Val Ala Asp Ala Ala Glu Ala Phe Arg Ala
130 135 140

Ser Arg Arg Arg Gly Ala Arg Pro Ala Phe Ala Pro Val Asp Leu Gly
145 150 155 160

```
Arg Gly Phe Ala Phe Ala Glu Val Glu Leu Tyr Gly Asp Val Val Leu
                165                 170                 175

Arg Phe Val Ser His Pro Asp Gly Thr Asp Val Pro Phe Leu Pro Gly
            180                 185                 190

Phe Glu Gly Val Thr Asn Pro Asp Ala Val Asp Tyr Gly Leu Thr Arg
        195                 200                 205

Phe Asp His Val Val Gly Asn Val Pro Glu Leu Ala Pro Ala Ala Ala
    210                 215                 220

Tyr Ile Ala Gly Phe Thr Gly Phe His Glu Phe Ala Glu Phe Thr Ala
225                 230                 235                 240

Glu Asp Val Gly Thr Thr Glu Ser Gly Leu Asn Ser Val Val Leu Ala
                245                 250                 255

Asn Asn Ser Glu Gly Val Leu Leu Pro Leu Asn Glu Pro Val His Gly
            260                 265                 270

Thr Lys Arg Arg Ser Gln Ile Gln Thr Phe Leu Glu His His Gly Gly
        275                 280                 285

Pro Gly Val Gln His Ile Ala Val Ala Ser Ser Asp Val Leu Arg Thr
    290                 295                 300

Leu Arg Lys Met Arg Ala Arg Ser Ala Met Gly Gly Phe Asp Phe Leu
305                 310                 315                 320

Pro Pro Pro Leu Pro Lys Tyr Tyr Glu Gly Val Arg Arg Leu Ala Gly
                325                 330                 335

Asp Val Leu Ser Glu Ala Gln Ile Lys Glu Cys Gln Glu Leu Gly Val
            340                 345                 350

Leu Val Asp Arg Asp Asp Gln Gly Val Leu Leu Gln Ile Phe Thr Lys
        355                 360                 365

Pro Val Gly Asp Arg Pro Thr Leu Phe Leu Glu Met Ile Gln Arg Ile
    370                 375                 380

Gly Cys Met Glu Lys Asp Glu Arg Gly Glu Glu Tyr Gln Lys Gly Gly
385                 390                 395                 400

Cys Gly Gly Phe Gly Lys Gly Asn Phe Ser Glu Leu Phe Lys Ser Ile
                405                 410                 415

Glu Asp Tyr Glu Lys Ser Leu Glu Ala Lys Gln Ser Ala Ala Val Gln
            420                 425                 430

Gly Ser
```